

Description

[Cobalt Silicidation Process for Substrates Comprised with a Silicon – Germanium Layer

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This patent application claims the benefit of Taiwanese patent application Serial No. filed on June 12, 2003.

BACKGROUND OF INVENTION

[0002] Performance of complimentary metal oxide semiconductor (CMOS) devices have been enhanced via increases in drive current achieved via strain induced band structure modification. Channel regions formed in a silicon layer under bi-axial tensile strain have allowed enhanced electron mobility to be realized for Nor NMOS devices. This is accomplished via formation of the silicon layer on an underlying relaxed semiconductor alloy layer, such as a silicon – germanium layer, which in turn is formed on the underlying semiconductor substrate. In addition, enhanced hole mo-

bility can be realized via formation of a Por PMOS device in a silicon – germanium layer with biaxial compressive strain. The presence of the performance enhancing silicon – germanium layer however presents difficulties during subsequent metal silicide formation processes. For example, silicidation processes, applied to a CMOS source/drain region will consume part or all of the silicon – germanium layer overlying the source/drain region. During the silicidation process, germanium atoms tend to segregate and accumulate at the surfaces of the metal silicide grains. The resulting germanium-rich interfacial layer at the surface of the metal silicide grains behave as a diffusion barrier and retard complete formation of the desired metal silicide layer, thus resulting in a lower than desired conductivity of the metal silicide layer when compared to metal silicide counterparts formed from non-semiconductor alloy layers. The decrease in metal silicide conductivity or the increase in metal silicide resistance will negatively influence CMOS performance.

BRIEF DESCRIPTION OF DRAWINGS

[0003] The object and other advantages of this invention are best described in the preferred embodiment with reference to the attached drawings that include: Figs. 1-12 are cross-

sectional views of first and second embodiments of a method used to partially or totally remove portions of a silicon – germanium layer from the surface of source/drain regions prior to formation of metal silicide on the source/drain region.

[0004] Figs. 13–18 are cross-sectional views of third and fourth embodiments of a method used to reduce unwanted germanium segregation in the grains of the metal silicide layer during the metal silicide formation procedure.

DETAILED DESCRIPTION

[0005] Various embodiments of a method for formation of a metal silicide layer over source/drain regions of a semiconductor device with a silicon – germanium layer are described herein. Embodiments provide for the removal (full or partial) or altering of silicon – germanium. It should be noted that the various process parameters described herein are provided as exemplary details and that other suitable procedures and parameters may be used to accomplish the same.

[0006] In Fig. 1, a semiconductor substrate 1 comprising single crystalline silicon featuring a <100> crystallographic orientation may be used. To form a P or PMOS device, a semiconductor alloy layer such as a silicon – germanium layer

2a, obtained under biaxial compressive strain, may be used to enhance hole mobility in a channel region located in the compressively strained silicon – germanium layer. The silicon – germanium layer 2a may be epitaxially grown on the semiconductor substrate 1, to a thickness between about 40 to 400 Angstroms, featuring a weight percent of germanium between about 10 to 40 percent, for example. To avoid partial consumption of silicon – germanium layer 2a during subsequent growth of an overlying gate insulator layer, a silicon capping layer 3, may be epitaxially grown over the silicon germanium layer 2a. The silicon capping layers may have a thickness between about 20 to 100 Angstroms.

[0007] If an NMOS or Ndevice is to be formed, enhanced electron mobility or enhanced performance may be accomplished by locating the NMOS channel region in a silicon layer 4 under biaxial tensile strain. This is schematically shown in Fig. The silicon layer 4 may be epitaxially grown over the underlying silicon – germanium layer 2a. The silicon layer 4 may have a thickness between about 20 to 100 Angstroms. Again to avoid consumption of the material used to accommodate the device channel region during gate insulator formation, a silicon capping layer 3, is

formed over the silicon layer 4. If desired, the process herein may also be applied to an NMOS device in which a silicon layer, under biaxial tensile strain, overlays a silicon – germanium layer.

[0008] In Fig. 3, a gate insulator layer 24, comprised of silicon dioxide may be thermally grown to a thickness between about 8 to 50 Angstroms, for example, over the silicon capping layer 3. A conductive layer such as a doped polysilicon layer 5, may be deposited via low pressure chemical vapor deposition (LPCVD) to a thickness between about 600 to 2000 Angstroms, for example. The polysilicon layer 5 may be doped in situ during deposition via the addition of arsine or phosphine to a silane ambient, or the polysilicon layer 5 may be deposited intrinsically then doped via implantation of arsenic or phosphorous ions. If lower line resistance is desired, a composite layer (not shown) comprised of an underlying doped polysilicon layer and an overlying metal silicide layer may be employed. The metal silicide layer can be a titanium silicide or tungsten silicide layer, for example, or other suitable materials. A photoresist may be used to pattern etch and define conductive gate structure from the polysilicon layer 5. An anisotropic reactive ion etching (RIE) procedure may be

used using chlorine or fluorine-based gas as a selective etchant for the polysilicon, or for the composite layer. Removal of the photoresist may be accomplished using known or later developed plasma oxygen ashing and wet clean procedures. A buffered hydrofluoric acid component of the wet clean procedure may remove a portion of the gate insulator layer 24 not covered by the polysilicon layer 5.

[0009] If desired, a lightly doped source/drain (LDD) region (not explicitly shown) may be formed in the semiconductor alloy layer 2a and in the semiconductor substrate 1 not covered by the conductive gate structure 5. The LDD regions may be formed by known processes such as ion implantation. The LDD region may also be defined in the exposed portion of the silicon capping layer 3. An insulator layer 6 such as silicon nitride is formed over the device. The insulator layer 6 may have thickness between about 150 to 1000 Angstroms formed using LPCVD or plasma enhanced chemical vapor deposition (PECVD), for example. An anisotropic RIE procedure, using CF_4 as an etchant for silicon nitride, for example, is employed to define the insulator layer spacers 6 on the sides of conductive gate structure 5. Heavily doped source/drain regions 7, are

formed via implantation of the desired species in portions of silicon – germanium layer 2a, and in portions of semiconductor substrate 1 not covered by the gate structure 5, or insulator spacers 6. Heavily doped source/drain regions 7 for a PMOS device may be formed via implantation of boron or BF_2 ions, for example. For an NMOS device, the heavily doped source/drain regions 7 may be formed via implantation of arsenic or phosphorous ions for example.

[0010] For performance optimization, minimum source/drain resistance is desired. This can be accomplished via formation of a metal silicide region on the underlying source/drain region. The metal silicide region may be formed via consumption of portions of the silicon – germanium layer 2a, as well as via consumption of portions of semiconductor substrate 1. However, during the silicidation procedure, germanium atoms may segregate and accumulate at the surface of the metal silicide grains. The germanium-rich interfacial layer at metal silicide surfaces may act as a diffusion barrier and retard complete formation of the metal silicide region, thus not allowing the desired minimum source/drain resistance to be realized. A first embodiment of this invention in which the portion of silicon – germanium layer 2a, is either partially or totally re-

moved from above the source/drain regions prior to the silicidation procedure, is now addressed and schematically shown in Figs. 5 – 10.

[0011] Referring to Fig. 5, partial removal of silicon – germanium layer 2a, is accomplished via a selective anisotropic RIE procedure, using Cl_2 , or SF_6 , or a fluorine-based gas as an etchant, for example. The remaining silicon germanium portion 2b may have a reduced thickness between about 20 to 200 Angstroms. The thinner silicon – germanium layer 2b may result in less germanium segregation at the surfaces of a metal silicide region during a silicidation procedure, when compared to counterpart metal silicide regions formed on thicker silicon – germanium layers. Total removal of exposed portions of silicon – germanium, shown schematically in Fig. may be accomplished via a more prolonged selective anisotropic RIE procedure. In other words, by varying the process time of the etching process, the silicon germanium layer 2a may be partially or totally removed depending on the desired result. In both examples the selectivity of the RIE procedure features the non – etching of insulator spacers 6, however the conductive gate structure 5 may be thinned during the anisotropic RIE procedure used to partially or totally re–

move silicon – germanium from the top surface of source/drain regions 7.

[0012] In Fig. 7, deposition of a metal layer such as cobalt, via a process such as physical vapor deposition (PVD) may be performed. An anneal procedure follows. The resulting metal silicide structures regions 8, are over the source/drain regions 7, as well as on the conductive gate structure 5. The anneal may be a rapid thermal anneal (RTA) process at a temperature between about 300 to 900°C, for between 0.1 to 180 seconds in an inert ambient, for example. The metal silicide regions 8 may have a thickness between about 50 to 150 Angstroms. If desired, the anneal process may be performed using conventional furnace procedures. The metal silicide regions 8, such as cobalt silicide regions may be formed via the consumption of the semiconductor substrate 1, for the case in which silicon – germanium was totally removed prior to the silicidation procedure, or formed via consumption of the thinned silicon – germanium layer 2b, in addition to the consumption of a top portion of semiconductor substrate 1, underlying the thinned silicon – germanium layer 2b. The metal silicide regions 8 may also be formed on the top surface of the gate structure 5, however portions of

the metal layer, cobalt in this example, located on the surface of insulator spacers 6, may remain unreacted. Unreacted cobalt may be selectively removed via a wet etch procedure such as by using a solution of HNO_3 or HF, for example.

[0013] A second embodiment of this invention is the partial or total removal of portions of a silicon – germanium layer located over the source/drain regions via oxidation of the exposed portions of silicon – germanium followed by the removal of the oxidized regions. Fig. 8 schematically shows the result of a thermal oxidation procedure used to form a silicon dioxide region 9b, by total consumption of the silicon – germanium layer 2a. This may be accomplished by a thermal oxidation procedure performed in an oxygen – steam ambient, for example. The thermal oxidation process temperature may be between about 500 to 1000°C resulting in the silicon oxide region 9b at a thickness between about 8 to 50 Angstroms, over the source/drain regions 7. If desired, the thermal oxidation procedure can be performed using conditions that will convert only a top portion of the silicon – germanium layer to silicon oxide, thus leaving a bottom portion of the silicon – germanium on the underlying source/drain regions 7. The

thermal oxidation procedure may also result in the formation of a silicon oxide region 9a, at a thickness between about 600 to 2000 Angstroms, on the gate structure 5.

[0014] The selective removal of the silicon oxide regions 9a and 9b is next addressed and schematically shown in Fig. 9. A wet etch procedure using a buffered hydrofluoric acid solution selectively removes the silicon oxide region 9b, a silicon oxide region formed via consumption of silicon – germanium, from the top surface of the source/drain regions 7. The selective wet etch procedure may also remove the silicon oxide region 9a, from the top surface of the conductive gate structure 5. If desired, the selective removal of the silicon oxide regions may be accomplished via dry etching procedures using CHF_3 , for example, as a selective etchant for silicon oxide. The total or partial removal of silicon – germanium via an oxidation – removal sequence allows a subsequent silicidation procedure to be performed either with or without a reduced level of germanium segregation at the surface of the metal silicide grains. The silicidation procedure, similar or identical to the procedure previously described is again performed to form a cobalt silicide region 8 on the source/drain regions 7, as well as on the conductive gate structure 5. This is

shown in Fig. 10.

[0015] In Fig. 11, an iteration of the second embodiment of this method entails partial or total removal of portions of silicon – germanium via formation and removal of disposable metal silicide regions, prior to the formation of the desired metal silicide regions on the source/drain region. The disposable metal silicide layers are formed via deposition of a cobalt or titanium layer 10a 10c via PVD procedures, for example. The thickness of the cobalt or titanium layer 10a 10c is chosen to either partially or totally remove silicon germanium from above the source and drain regions. For this example, the total consumption of silicon – germanium overlying source/drain regions 7, in semiconductor substrate I may comprise depositing cobalt or titanium layer 10a, to a thickness between about 50 to 150 Angstroms. A first RTA procedure may be performed, at a temperature between about 300 to 900°C, for a time between about 0.1 to 180 seconds, resulting in the formation of disposable metal silicide region 10c (cobalt silicide or titanium silicide) on the source/drain regions 7, and metal silicide region 10b on the gate structure 5. Portions of metal layer 10a located on the insulator spacers 6 may remain unreacted. If only partial removal of

silicon – germanium is desired, the disposable metal silicide region 10c would overlay an unconsumed bottom portion of the silicon – germanium layer.

[0016] A selective wet etch procedure may be used to remove the unreacted portions of metal layer 10a from the surface of the insulator spacers 6, followed by a selective dry or wet etch procedure used to remove the disposable metal silicide regions 10c and 10b from the source/drain and conductive gate regions. The selective dry etch option is accomplished using Cl_2 , SF_6 or Fluorine-based gases, for example, as a selective etchant for the disposable metal silicide regions, terminating at the appearance of underlying source/drain regions 7, and underlying conductive gate structure 5. The selective wet etch option may be performed using HNO_3 , HF or H_3PO_4 , for example, as the selective etchant for both the disposable metal silicide region 10c on the source/drain regions 7 and the disposable metal silicide region 10b located on the conductive gate structure 5. The formation of the desired metal silicide region on both source/drain regions 7 and on conductive gate structure 5 may be accomplished via the deposition of cobalt via PVD procedures, for example, at a thickness between about 50 to 150 Angstroms. A second

RTA procedure may be employed at a temperature between about 300 to 900°C, for a time between about 0.1 to 180 seconds, in an inert ambient, resulting in the formation of metal silicide, or cobalt silicide region 8, on source/drain regions 7, and on conductive gate structure 5, as shown in Fig. 12. The cobalt silicide region 8 located on source/drain regions 7 is without germanium at the grain surfaces thus allowing the complete formation of the low resistance cobalt silicide region. If partial removal of silicon – germanium were accomplished via the formation and removal of a disposable metal silicide region, the subsequent cobalt silicide region would still be formed with a reduced level of germanium segregation when compared to counterparts formed consuming the entire thickness of underlying silicon – germanium. The removal of portions of unreacted cobalt located on the surface of insulator spacers 6 may be selectively accomplished via wet etch procedures, again using a solution of HNO_3 , HF or H_3PO_4 , for example.

[0017] Additional embodiments of the method described herein are described and schematically shown in Figs. 13 – 15. A cobalt layer is deposited via PVD procedures, for example, to a thickness between about 50 to 100 Angstroms. A first

RTA procedure may be next employed, at a temperature between about 300 to 900°C for a time between about 0.1 to 180 seconds in an inert ambient, resulting in the formation of a $\text{Co}(\text{Si}_x\text{Ge}_{1-x})$ layer 11b on a bottom portion of silicon germanium layer 2a, in a region in which silicon – germanium layer 2a, overlays the source/drain regions 7. The RTA procedure also results in metal silicide, or cobalt silicide layer 11a, on the conductive gate structure 5, while portions of the cobalt layer located on insulator spacers 6, remain unreacted. The selective removal of unreacted cobalt may be accomplished by the use of wet etch procedures, for example. The result of these procedures is schematically shown in Fig. 13.

[0018] To reduce the segregation of germanium at the surface of cobalt silicide grains during a second RTA procedure, specific species may be first implanted into the $\text{Co}(\text{Si}_x\text{Ge}_{1-x})$ layer 11b. For example, a first group of implanted species may comprise nitrogen or silicon ions, implanted at an energy between about 10 to 100 KeV, at a heavy dose between about 10^{13} to 10^{16} atoms/cm². A second group of implanted species, offering the same retardation of germanium atoms may comprise arsenic or phosphorous ions, implanted at an energy between about 10 to

100 KeV, at a dose between about 10^{13} to 10^{16} atoms/cm². The second group may be implanted at a lighter dose than that of the first group of implanted species. Fig. 14 schematically shows implanted species 12, in both the $\text{Co}(\text{Si}_x\text{Ge}_{1-x})$ layer 11b over the source/drain 7 and the cobalt silicide layer 11a on conductive gate structure 5.

[0019] The second RTA procedure is now used to form a cobalt silicide region 13 on the source/drain regions 7. The second RTA procedure may be performed at a temperature between about 500 to 900°C, for a time between about 0.1 to 180 seconds, in an inert ambient. The presence of implanted species 12 in the $\text{Co}(\text{Si}_x\text{Ge}_{1-x})$ layer 11b reduces the segregation of germanium at the surfaces of the grains of the cobalt silicide region 13, allowing a minimum resistance of the metal silicide region to be realized. Cobalt silicide region 11a located on conductive gate structure 5, also comprised with implanted species 12, remains unchanged during the second RTA procedure. The result of the second RTA procedure is schematically shown in Fig. 15. Although this embodiment is described where only a top portion of the silicon – germanium layer 2a was converted to a $\text{Co}(\text{Si}_x\text{Ge}_{1-x})$ layer during the first RTA procedure, the entire portion of silicon – germanium

may be converted to $\text{Co}(\text{Si}_x\text{Ge}_{1-x})$ during the first RTA procedure and prior to the implantation procedure, if desired. The implanted species in the thicker $\text{Co}(\text{Si}_x\text{Ge}_{1-x})$ layer is still operable to reduce germanium segregation during the second RTA procedure.

[0020] A further embodiment of the method disclosed herein entails the implantation of the same species described above however prior to the first RTA procedure. Referring to Fig. 16, after the deposition of the cobalt layer 10a, the implantation procedure is performed and then followed by the first RTA procedure, resulting in a $\text{Co}(\text{Si}_x\text{Ge}_{1-x})$ layer 10c on the source/drain regions 7. The above procedures also results in the formation of cobalt silicide layer 10b on the conductive gate structure 5. The implantation of species 12 such as silicon ions may be performed at an energy between about 10 to 100 KeV, at a dose between about 10^{13} to 10^{16} atoms/cm², for example. The first RTA procedure may be performed at a temperature between about 300 to 900°C, for a time between about 0.1 to 180 seconds. The cobalt layer 10a residing on insulator spacers 6 remains unreacted during the first RTA procedure and may be selectively removed by the use of a solution comprised of HNO_3 , HF or H_3PO_4 , for example. This is

schematically shown in Fig. 17.

[0021] A second RTA procedure may be performed. The RTA process may use a temperature between about 500 to 900°C, for a time between about 0.1 to 180 seconds in an inert ambient, resulting in the formation of the cobalt silicide region 8 on source/drain regions 7. The presence of implanted species 12 in the $\text{Co}(\text{Si}_x\text{Ge}_{1-x})$ layer 10c minimizes germanium segregation to the grains of cobalt silicide region 8, resulting in minimum resistance in the metal silicide region. The result of the second RTA procedure is schematically shown in Fig. 18.

[0022] While this invention has been particularly shown and described with reference to, the preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made without departing from the spirit and scope of this invention.